论文的相关参数与被引频次的关系研究

肖学斌[1]　柴艳菊[2]
[1](武汉大学图书馆　武汉　430079)
[2](中国科学院测量与地球物理研究所　武汉　430077)

摘要:【目的】研究论文的某些相关参数是否会影响被引频次。【方法】采用多种措施弱化非研究因素的干扰，并绘制研究因素与被引频次关系的各年代曲线，判断研究因素对被引频次的影响。【结果】作者人数、页码数、参考文献数和文摘长度与被引频次呈正相关，作者关键词数量和平均长度与被引频次无关，而不同的题名长度对被引频次的影响不同。【局限】由于数据采样条件限制，数据均采自 SCIE 这种高水平的论文数据库，且 WOS 分类为 Engineering 和 Mechanical，所得结论不一定全部适用于其他主题的论文。【结论】论文的某些相关参数对论文的被引频次存在影响。

关键词: 被引频次　影响因素　字段长度　参考文献数　作者数　正相关
分类号: G237.5

1　引　言

在我国，论文的被引频次已经成为衡量论文和作者影响力的一个重要指标，日益受到评审机构和研究人员的重视，也日益受到期刊编辑部的重视。如何提高论文的被引频次是研究人员和期刊编辑普遍关注的问题。要回答这个问题，必须首先搞清楚有哪些因素会影响论文的被引频次，这些因素是如何影响被引频次的。关于这个问题，已有研究人员对部分因素做了研究。作者撰写论文时，由于受自身的学科知识背景、学识、研究条件和研究水平等制约，使得论文被接受的期刊及期刊相关因素(如期刊出版周期、发文量、研究领域等)、被收录的数据库等因素也受到制约，这些因素都可能会对被引频次产生影响。已有的研究认为:

(1) 文献的被引率与文献的体裁相关。综述和述评被引明显高于应用和实验研究类[1-2];

(2) 文献的被引率与研究领域(学科)有关[1-5]。热门的研究领域被引率高于冷门研究领域[3-4];

(3) 文献的被引频次受期刊相关因素的影响。期刊的影响因子越高，文献被引率越大[6]。期刊的出版周期越长，发文量越高，文献被引率越大[3]；期刊印刷质量越好，文献被引率越高[7];

(4) 编辑部对篇幅的规定，编辑的工作态度对文献被引率有一定的影响[3,5,7];

(5) 收录期刊的数据库知名度越高，被引率越高[5,8]；数据库分布的范围越广，文献被引率越高[7];

(6) 论文存在的形式影响被引率。论文的获取和传播越容易，被引率越高[5,9]。

还有一些论文本身的相关参数，如标题的长度、关键词的个数与关键词的长度、文摘的长度、论文的长度、参考文献数等，这些因素是否也会影响论文的被引频次呢？本文重点研究这些相关参数对被引频次的影响。其中有些虽有文献研究，但是不同文献得到的结论不一致。如参考文献[1]研究结论是“4-6 个作者的论文被引率和篇均被引频次最高，作者数多于 7 个的论文，篇均被引频次反而低于 4-6 个作者的论文”；而参考文献[2]的结论是“期刊的作者数量与影响因子成正比”。参考文献[9]认为：论文标题长度和论文长度对被引频次没有明显的影响，关键词数量对被引存在明显的影响。而参考文献[10]却认为论文的长度对被引频次存在影响。笔者也对此做了研究。下面介绍笔者的研究思路和研究方法。

通讯作者：肖学斌，ORCID：0000-0003-1933-5006，E-mail：00201493@whu.edu.cn。

2 相关参数对被引频次的影响研究

2.1 凸显特定因素影响到措施

由于同时影响论文被引频次的因素很多，为了研究其中一个因素与被引频次的关系，需要采取许多措施来弱化其他因素和相关参数的影响，凸显特定因素的影响。措施如下：

(1) 限定所收录数据库。由于数据库对论文的被引频次有影响[5,8]，所以限定数据库，保持文献所收录数据库的一致性，可以弱化数据库对论文被引频次的影响。笔者只选用 SCIE 数据库作为采样数据源。

(2) 保持文献水平的一致性。由于所采样数据都来源于SCIE这一国际公认的，代表高水平研究论文的数据库，可以保证所选论文有较多的文献被引，避免由于被引论文数量过少导致结果不明显或者失真。SCI 论文的被引情况也是学术评价关注的重点之一。

(3) 限定主题。由于主题研究领域对被引频次有较大影响[1-5]，笔者于 2015 年 9 月 25 日在 SCIE 数据库中检索 2010 年–2012 年 WOS 分类包含 Engineering 和 Mechanical 的全部文献，检索得到 46 378 条记录，从而保证检索出的文献主题研究领域比较接近，弱化由于主题研究领域不同造成的被引频次的差异。

(4) 限定出版物类型。在 SCIE 数据库中所收录的出版物类型有期刊、图书、丛书、专利 4 种类型，为了防止出版物类型可能对被引频次造成的影响，滤掉非期刊类型的记录 9 条，再去掉 6 条信息不全的记录，最终得到 46 363 条期刊论文记录。

(5) 大数据量。众所周知，取大数据量的平均值可以弱化偶然因素的影响，凸显出主要趋势，显现普遍规律。笔者样本数高达 46 363。而参考文献[1-10]中最大样本数为 5 716[2]，最少的只有几十个，大多数总样本数不足 1 000。

(6) 移动平均法。由于影响论文被引频次的因素很多，尽管采用了上面的种种措施，依然不能完全消除其影响，导致某些地方波动幅度过大，因此有时有必要利用移动平均法弱化异常波动以保持总体变化趋势。

2.2 研究及判断方法

首先将原记录导入 Excel 中，运用 VBA 提取原始记录表中的被引频次、发表年代、页码数、作者数、参考文献数、基金支持等字段数据保存于另一张数据表中，计算题名长度(单词数)、文摘长度(单词数)、关键词数量、关键词长度等数据，也存放于上面的数据表中。并计算各种情形下的平均被引频次，其中，由于文摘长度跨度、参考文献数跨度巨大，许多具体情况对应的数量可能会比较少，甚至为 0，故取几个相邻量的平均值来进行分析。

然后对数据进行排序，凸显被研究因素。例如，在分析论文长度(用页码数表示)与被引频次关系的时候，由于论文的被引频次会随时间累积，因此分年统计，先按照出版年代升序排列，将不同年代的论文分开；接着按照页码数(相关参数)升序排列以凸显页码数的影响效果。以页码数作为自变量，被引频次作为因变量，绘制关系图，得到三条平均被引频次关于论文页码数的曲线。如果这三条曲线图形基本一致，则表示该趋势具备普遍性，否则可能出现了错误；如果三条曲线总趋势(上升、下降)基本一致，则表示该因素对论文被引频次有影响；如果三条曲线总趋势都是水平的则没有影响；如果三条曲线一致，但是各阶段趋势不同，则表示不同阶段自变量对被引频次影响不同。这是判断相关参数是否影响被引频次的依据。笔者采用这种方法，研究了页码数、作者数量、题名长度、文摘长度、作者关键词数量及平均长度、参考文献数和是否基金支持等因素与平均被引频次的关系，并根据计算结果绘制成图，如图 1–图 8 所示：

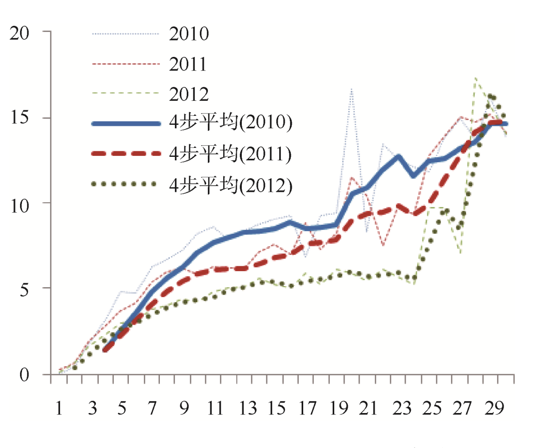


图 1 页码数与被引频次的关系

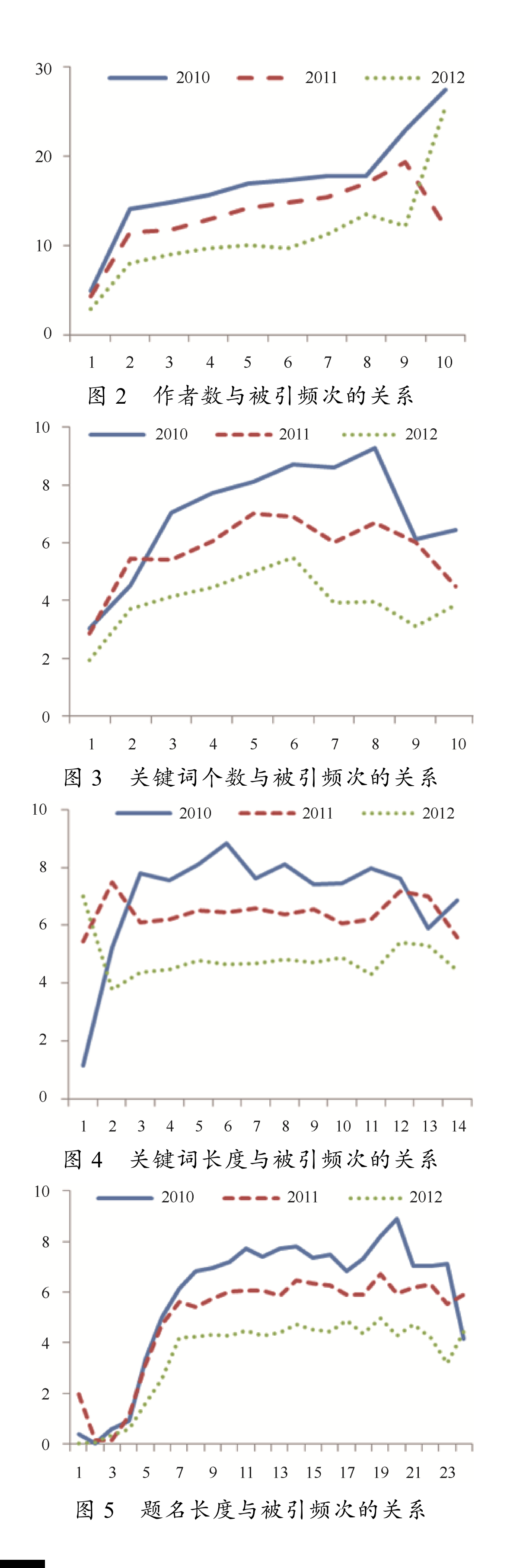


图 2　作者数与被引频次的关系

图 3　关键词个数与被引频次的关系

图 4　关键词长度与被引频次的关系

图 5　题名长度与被引频次的关系

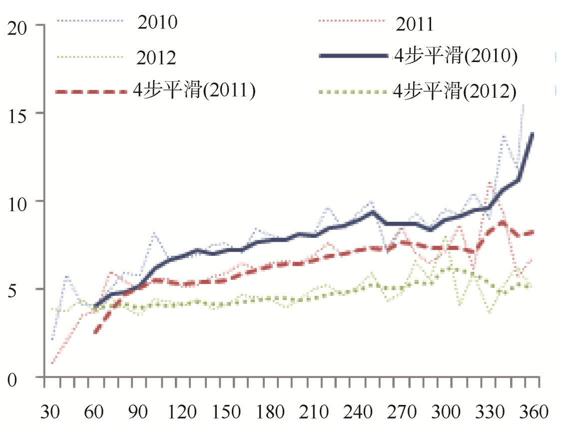


图 6　文摘长度与被引频次的关系

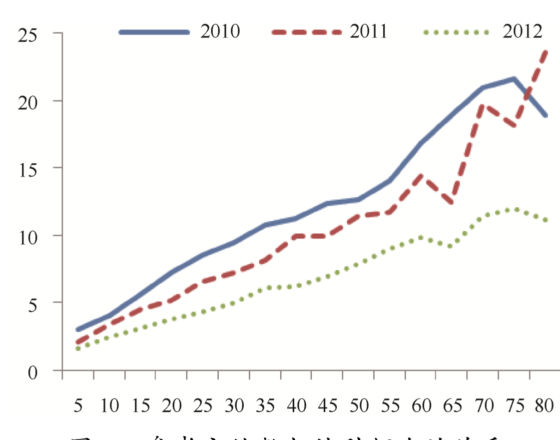


图 7　参考文献数与被引频次的关系

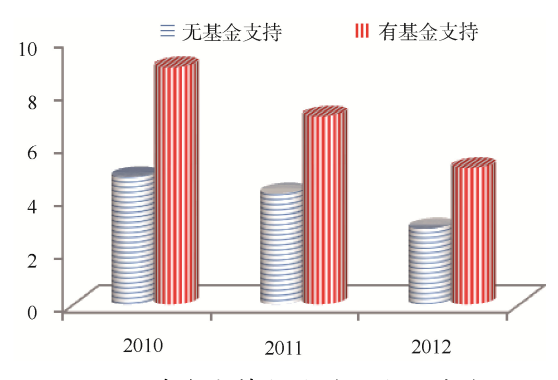


图 8　基金支持与被引频次的关系

2.3　相关性分析

为了确定平均被引频次与前面所述相关参数之间是否确实存在关系，笔者调用 Excel 的相关性函数 Correl 计算其间的相关性，包括与各年平均被引频次的相关性，与全部平均被引频次的相关性，为了减少偶然因素的影响，特地分析了子样本数大于10的情况下与平均被引频次的相关性，相关性检验结果如表 1 所示。

表 1　相关参数与平均被引频次的相关度

相关参数	全部论文				子样本数大于 10				
	2010	2011	2012	三年	2010	2011	2012	三年	相关度 r
页码数	0.754	0.793	0.872	0.797	0.919	0.954	0.838	0.950	非常高
作者数	0.810	0.799	0.807	0.649	0.900	0.727	0.823	0.924	非常高
关键词数	0.520	0.363	0.213	0.461	0.520	0.363	0.213	0.461	低
关键词长度	0.189	0.189	–0.214	0.487	0.189	0.189	–0.214	0.487	低
题名长度	0.567	0.297	0.434	0.447	0.678	0.687	0.566	0.447	低
题名长度≤8	0.955	0.862	0.952	0.935	0.955	0.862	0.952	0.935	非常高
文摘长度	0.688	0.816	0.536	0.822	0.628	0.803	0.467	0.808	高
参考文献数	0.932	0.954	0.977	0.986	0.906	0.957	0.977	0.986	非常高

在相关分析中，一般根据相关度r的数值大小，将不完全线性相关的密切程度分为4个等级：0＜|r|≤0.3，微弱相关；0.3＜|r|≤0.5，低度相关；0.5＜|r|≤0.8，中度相关；0.8＜|r|≤1，高度相关[11]。据此标准，在表1中，当参与统计论文数大于10时，有4个相关参数与被引频次的相关性非常高，相关度大于0.92，分别是页码数、作者数、参考文献数和题名长度≤8时的题名长度，文摘长度也与被引频次有较高的相关性，相关度大于0.8，而关键词数、关键词长度和全部题名长度的相关性低，均小于0.5。如果用全部论文数时计算，只有题名长度≤8时和参考文献数与被引频次的相关性非常高，分别是0.935和0.986；页码数、作者数和文摘长度与被引频次中度相关，相关度在0.649-0.822之间。故可以得出结论，页码数、作者数、参考文献数以及题名长度≤8对被引频次有明显的影响。两种方式之间相关性差异主要是由于偶然因素造成的。

2.4　数据分析结果及结论

从总体上看，图 1–图 8 中，2010 年的平均被引频次都是最高，2012 年的都是最低，符合被引频次随时间累积的规律；而且图 1–图 7 中各自的三条曲线都大体相似，图 8 中基金支持与非基金支持的论文平均被引频次比也大致相等，分别为 1.86、1.70、1.80，都说明这些图确实反映了论文平均被引频次与这几种相关参数的关系。虽然每幅图各自的三条曲线的趋势相似，但是不同图的曲线却有差异(包括倾斜度和形状)。将图 1–图 7 和表 1 结合起来，得出结论如表 2 所示：

表 2　论文相关参数对被引频次影响

相关参数	图形曲线趋势	相关性	结论
页码数	明显上升	非常强	被引频次与论文长度密切正相关
作者数	逐渐上升	非常强	被引频次与作者数密切正相关
关键词数	无明显趋势	弱	关键词数与被引频次无关
关键词长度	无明显趋势	弱	关键词长度与被引频次无关
题名长度	无固定趋势	弱	题名长度对被引频次影响不确定
题名长度≤8	逐渐上升	非常强	题名长度≤8 与被引频次密切正相关
文摘长度	缓慢上升	强	文摘长度与被引频次正相关
参考文献数	缓慢上升	非常强	被引频次与参考文献数密切正相关

图 8 显示受基金支持的论文被引频次更高。2010 年–2012 年有基金支持的论文数分别是 6 368，7 502，8 680；无基金支持的论文数分别是 8 252，7 946，7 615。这三年的数据正好代表了三种情况：2010 年有基金支持的论文数明显低于无基金支持的论文数；2011 年两者基本持平；2012 年则明显高于无基金支持的论文数(基金支持比例逐年提高，是否表示 SCIE 论文更青睐于基金支持论文？有待研究)。但是无论哪种情况，有基金支持的论文平均被引频次都明显高于无基金支持的论文的被引频次。这表明基金支持对文献的被引频次有明显的影响，与文献[12-14]结论一致。

2.5　相关参数与被引频次的关系理论分析

不可否认，一篇论文是否会被引用，引用者必须经历论文的发现、获取、阅读和引用4个阶段(学术造假除外)。被发现、被获取和被阅读是被引用的先决条件，也就是说，论文越容易被发现，被引用的可能性就越大；论文全文越容易被获取，被引用的可能性也越大[5,9-10]。网络时代，利用网络查找文献的方式可分为两种：直接利用搜索引擎查找和获取，这是大多数人采用的方式，因为其便宜、快捷、门槛低[15]；利用专业文献数据库查找，由于专业文献数据库使用费用比较昂贵，一般高校每年会花费几百万到几千万元的经费购买数据库的使用权，而且使用专业文献数据库需要学习相关的检索知识，因此这种方式只有那些拥有这些专业文献数据库使用权限的单位和个人才可能采用。搜索引擎查找一般是模糊匹配方式，即搜索引擎会将检索词拆散，甚至拆成单字或单词，然后在搜索引擎数据库中进行匹配，并将匹配结果根据相关度从高到低进行排序[16]。从这个理论出发，很容易解释被引频次与文献长度、文摘长度正相关的关系。研究人员通常会用几个检索词来查找所需要的文献，而全文和文摘比较长，检索词可以出现在不同的句子中，全文越长、文摘越长，被命中的几率就越大，因而被引用的可能性就更大。

可能有人会将这个结论推广到论文题名，从而认为：题名越长，被引频次越高。但是，图5显示的被引频次与题名长度的关系并非如此。2010年–2012年的曲线均是：当题名长度小于8时，被引频次随题名的长度快速增加；当题名长度在8-20之间，被引频次变化缓慢；当题名长度超过20之后，被引频次反而随着题名长度的增加而降低。由于三条曲线情况基本一致，说明这种变化不是偶然，而是必然，究竟是什么原因造成的呢？为此，笔者做了专门研究。

首先针对题名单词数大于20的记录，提取被引频次、题名本身和单词数，存放于一个单独的Excel工作表中，得到1 378条记录。按照被引频次从低到高排序，发现很多论文是对另外一篇论文的评述(以Comments on开头)、回复(以Reply to或Response to开头)、讨论(以Discussion of开头)或者收回论文的申明(以Closure to开头)等，这些题名中除了包含另一篇论文的题名外，还包含其作者、刊名、年卷期页码等信息，从而导致论文题名很长。这样的论文有104篇，平均单词数约为26，共计被引50次，平均被引频次仅为0.48，其中有79篇被引频次为0，约占75.96%。这些论文都有很强的针对性，对大多数其他作者的研究意义不大，从而导致被引频次低。

在去掉这些论文后，计算结果显示：当题名长度超过20后，随着题名长度的增加，被引频次总体上依然逐渐降低。笔者根据上述104篇论文的被引情况猜测是由于论文的专指度太高导致的。一篇论文被发现后，是否阅读，最终还处决于引用者的研究兴趣与被引论文所涉及的研究内容的相关程度。虽然题名越长，文献检索时被命中的几率越大；但同时，题名越长，题名所包含的独立概念越多，表示所研究的内容越具体，研究范围越窄。如图9所示，A、B、C各表示论文题名中涉及一个独立概念的文献范围，独立的意思是三个概念相互之间不存在隶属关系；E(中间阴影区域)表示同时包含A、B、C三概念的文献范围。可以看出，E区域所在范围比任何一个概念所占区域小得多。科技论文的标题通常只有一个句子，是对论文表达内容的高度概括，全面或从不同的侧面体现作者的写作意图、研究主题或体现论文的亮点。如果题名比较长，研究人员可能会因为其中的某些概念与自己的研究兴趣相去较远而忽视这篇论文，自然不会引用；题名越长，独立概念越多，被忽视的可能性越大。因此，论文题名过长，反而会导致被引频次降低。被命中的概率和被关注的概率，这两个因素同时影响着被引频次，题名越长，被命中的概率越大，同时被关注的概率越低。图5可能是这两种因素的影响叠加的结果。

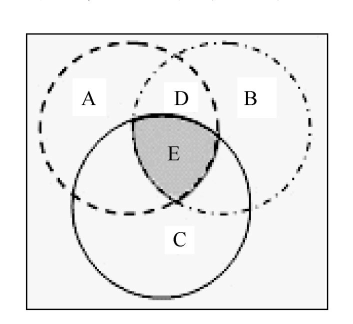


图9　多独立概念交叉范围E

图2显示论文的平均被引频次随着作者的人数增加而增加。一篇论文的多个作者必然有共同的研究兴趣，在后续的研究中，通常会相互关注其他成员的研

究成果，因此团队成员间研究成果的引用几率呈几何倍数增加。同时，由于各作者还可能有自己的研究团队，各作者自己的团队成员也可能会关注该论文，也会导致论文的被引频次增加。

图 3 和图 4 显示的论文作者关键词个数以及长度均对平均被引频次没有明显的影响。参考文献[15]研究表明：在检索文献时，多达 90%的大学生经常使用搜索引擎，仅有约 37%的大学生使用过中国期刊全文数据库。这就是说，搜索引擎是大学生查找参考资料的首选工具。大学生有免费文献数据库的使用权限，许多学生还学习过文献检索课，懂得数据库的使用方法，他们尚且以搜索引擎为主，其他无法使用专业数据库的人则别无选择。在使用搜索引擎检索文献时，大多数检索词实际上是自由词，甚至是自然语言，而且大多数作者关键词数量比较少(通常在 5 个左右)，能够同时匹配几个检索词的几率非常低，故作者关键词对被引频次影响不明显。

图 7 显示了参考文献数与被引频次成正相关的关系。总体上说，参考文献数越多，文献的被引频次越高。因为参考文献越多，在研究时用于查找、阅读、学习所花费的时间越长，掌握的资料和理论更全面、更准确，层次更深，得出的结论更可靠，因而论文质量更高，被引用的可能性越大[6]；同时，由于该论文与参考文献具有较强的相关性，会受到其参考文献作者的关注。参考文献越多，受到的关注越多，也增大了被引可能性。另外，检索参考文献也是用来查找相关研究资料的一种重要途径。

3 研究存在的问题和局限性

笔者选用 WOS 中的 SCIE 数据库作为数据来源，一方面是为了弱化某些因素的影响；另一方面也为了获取的数据更有意义，因为在国内 SCI 论文及其被引情况受到国内学术界和科技部门广泛的关注，且数据采样比其他数据库方便，却也大费周折。由于笔者有限的数据采样条件，可能带来研究的局限性，列举如下：

(1) SCIE 论文是高水平论文的代表，那么低水平论文的被引情况未得到证实，因此本研究结论主要适用于高水平论文。

(2) 由于研究数据属于 WOS 的 Engineering, Mechanical 主题，并非全部主题，故所得到的结论不一定全部适合其他主题分类。

(3) SCIE 数据库是一个英文文摘型数据库，即使原始文献语种非英语，在 SCIE 中都是以英文表示。由于语种的差异，导致题名、文摘等长度计数可能存在差异。另外由于 SCIE 中偶尔有数据不完备的情况存在，如，某些论文题名为“Untitle”，有些作者为空等，也或多或少地影响计算结果。

(4) 这些相关参数本身在各数据段的分布并不均匀。2010 年题名长度为 11 的论文数就有 1 413 篇，而三年的全部数据中，题名长度超过 20 的论文总数只有 1 378 条，分布跨度约占全部的 40%(题名最小长度为 1，最大为 52)，数据量却只有不到全部数据的 3%，这种不均匀分布也导致平均值受到偶然因素的影响程度有差异，数量越少，受影响的程度越高，这也是笔者特地针对论文数大于 10 的情况做统计分析的原因。尽管如此，也不能完全消除这种由于数量少带来的影响，图 1–图 7 两端异常情况或许就源于此。

4 结 语

影响文献被引频次的因素包括很多，在相关参数中，论文长度、文摘长度、参考文献数和作者数量以及基金支持对文献的被引频次产生正相关影响；题名长度也会影响被引频次，但是其影响变化趋势随长度不同而不同。由于笔者的研究结论都是基于 SCIE 数据库中的 Engineering 和 Mechanical 主题得到的，因此是否全部适用于其他论文还有待进一步研究。

参考文献：

[1] 刘雪立，徐刚珍，方红玲，等. 如何提高医学期刊的影响因子——从《眼科新进展》论文分类被引情况谈起[J]. 中国科技期刊研究，2008，19(4)：659-661. (Liu Xueli, Xu Gangzhen, Fang Hongling, et al. How to Improve the Impact Factor of a Medical Journal: To Refer from the Classification Citation of Recent Advances in Ophthalmology[J]. Chinese Journal of Scientific and Technical Periodicals, 2008, 19(4): 659-661.)

[2] 陈淑娴. 期刊影响因子的影响因素分析[J]. 山东理工大学学报：社会科学版，2006，22(4)：110-112. (Chen Shuxian. An Analysis of Affecting Factor of Journal's Influencing Multiplier [J]. Journal of Shandong University of Technology: Social Sciences Edition, 2006, 22(4): 110-112.)

[3] 黄萍，罗彦卿，王艳，等. 科技期刊影响因子的影响因素

分析及提高措施[J]. 编辑学报, 2006, 18(S1): 180-181. (Huang Ping, Luo Yanqing, Wang Yan, et al. Affecting Factor Analysis of Journal of Science and Technology and Improving Measurement [J]. Acta Editologica, 2006, 18(S1): 180-181.)

[4] 陈家顺. 学术期刊"影响因子"的非学术因素分析[J]. 湖北师范学院学报: 哲学社会科学版, 2005, 25(5): 133-135. (Chen Jiashun. An Analysis of the Non-learned Factors in "Affecting Factors" of Learned Journals [J]. Journal of Hubei Normal University: Philosophy and Social Sciences, 2005, 25(5): 133-135.)

[5] 胡剑胜. 关于影响因子主要影响因素的讨论[J]. 安徽农业科学, 2009, 37(13): 封 2-封 3. (Hu Jiansheng. Discussion on the Main Influence Factors of Impact Factor [J]. Journal of Anhui Agricultural Sciences, 2009, 37(13).)

[6] 孙书军, 朱全娥. 内容质量决定论文的被引频次[J]. 编辑学报, 2010, 22(2): 141-143. (Sun Shujun, Zhu Quan'e. Quality of Contents Determines the Total Cites of an Article [J]. Acta Editologica, 2010, 22(2): 141-143.)

[7] 俞立平, 潘云涛, 武夷山. 基于分位数回归的期刊影响因子影响因素研究[J]. 图书情报工作, 2010, 54(16): 145-148. (Yu Liping, Pan Yuntao, Wu Yishan. Study on Influences to Academic Journal Impact Factor Based on Quantile Regression [J]. Library and Information Service, 2010, 54(16):145-148.)

[8] 肖红, 袁飞, 邬建国. 论文引用率影响因素——中外生态学期刊比较[J]. 应用生态学报, 2009, 20(5): 1253-1262. (Xiao Hong, Yuan Fei, Wu Jianguo. Factors Affecting Citations: A Comparison Between Chinese and English Journals in Ecology [J]. Chinese Journal of Applied Ecology, 2009, 20(5):1253-1262.)

[9] 简琳, 何静, 周剑. 论文被引的文本因素分析:多学科视角[J]. 图书情报工作, 2011, 55(20): 32-35. (Jian Lin, He Jing, Zhou Jian. Document Factors Impacting on the Citation of an Article: Multi-fields View [J]. Library and Information Service, 2011, 55(20): 32-35.)

[10] Hudson J. Be Known by the Company You Keep: Citations-Quality or Chance? [J]. Scientometrics, 2007, 71(2): 231-238.

[11] 蔡智澄, 何立民. 相关性分析原理在图书情报分析中的应用[J]. 现代情报, 2006, 26(5): 151-152. (Cai Zhicheng, He Limin. Application of Correlation Analysis Theory in Library and Information Analysis[J]. Modern Information, 2006, 26(5): 151-152.)

[12] 刘划, 徐国艳. 2009-2013 年国内航空领域高被引论文分析[J]. 电大理工, 2014(2): 68-69. (Liu Hua, Xu Guoyan. Analysis of High Cited Papers from 2009 to 2013 in the Field of Domestic Aviation [J]. Study of Science and Engineering at RTVU, 2014(2): 68-69.)

[13] 刘划, 齐飞. 2001—2010 年国内互联网学科领域的高被引论文分析[J]. 科技情报开发与经济, 2014, 24(15): 123-125. (Liu Hua, Qi Fei. Analysis of Highly-cited Papers of Internet Subject Area Published in Domestic Journals from 2001 to 2010 [J]. Sci-Tech Information Development & Economy, 2014, 24(15): 123-125.)

[14] 赵静, 接雅俐, 杜志波. 基金资助论文与无基金资助论文被引情况分析: 以部分高水平医学期刊为例[J]. 南京医科大学学报: 社会科学版, 2012(6): 499-501. (Zhao Jing, Jie Yali,Du Zhibo. Analysis of Differences in Citation Between Funded Papers and Non-funded Papers in Some High-level Medical Journals [J]. Acta Universitatis Medicinalis Nanjing: Social Science, 2012, (6): 499-501)

[15] 杨冬艳. 浅谈大学生电子文献检索中的常见问题[J]. 图书馆论坛, 2009, 29(5): 147-148. (Yang Dongyan. On the Common Problems of Electronic Documents Retrieval for University Students [J]. Library Tribune, 2009, 29(5): 147-148.)

[16] 搜索引擎相关度算法分析[EB/OL]. [2015-03-23]. http://wenku.baidu.com/link? url=z2nvKb9aBvywcMWYUs2efiqigbjyeJCkW-pwom9CMToOMar_a0xqsiSLgH1ID6ySlCDPhKoGcbF7orn16uVV2ZMipjlTlE2TpjngfVT5gam. (Search Engine Correlation Algorithm Analysis [EB/OL]. [2015-03-23]. http://wenku.baidu.com/link?url=z2nvKb9aBvywcMWYUs2efiqigbjyeJCkW-pwom9CMToOMar_a0xqsiSLgH1ID6ySlCDPhKoGcbF7orn16uVV2ZMipjlTlE2TpjngfVT5gam.)

作者贡献声明:

肖学斌: 提出研究命题, 数据采样及分析, 论文起草及修改;
柴艳菊: 相关度验证方案, 修改论文。

利益冲突声明:

所有作者声明不存在利益冲突关系。

支撑数据:

支撑数据由作者自存储, E-mail: xxb@lib.whu.edu.cn。

[1] 肖学斌. 机械工程. xlsm. 论文的相关参数与被引频次的关系研究. 数据下载地址: http://pan.baidu.com/s/1i4P4REt.

收稿日期: 2016-01-12
收修改稿日期: 2016-03-02

Properties of Scholarly Papers and Number of Citations

Xiao Xuebin[1]　Chai Yanju[2]
[1] (Wuhan University Library, Wuhan 430079, China)
[2] (Institute of Geodesy and Geophysics, Chinese Academy of Sciences, Wuhan 430077, China)

Abstract: [**Objective**] To examine the ties between properties of scholarly papers and the number of citations they received. [**Methods**] First, we adopted various measurements to reduce the influence of irrelevant factors. Second, we drew trending lines to analyze the relationship between the target properties and the number of citations for a period of three years. [**Results**] There was positive correlation between some properties, such as the numbers of authors, pages and references, as well as the length of abstract, and the number of citations. In the meantime, there is no relationship between the number of keywords and the number of citations. The titles posed mixed effects to the number of citations. [**Limitations**] All samples were collected from the SCIE database in the fields of Engineering and Mechanical. We might not be able to get similar results from other areas. [**Conclusions**] Specific properties of the paper pose positive effects to the number of citations.

Keywords: Citations　Impact factors　Field length　The number of references　The number of authors　Positive correlation

---

Ex Libris 和 CLA 合作促进学生获取数字化内容

ProQuest 子公司 Ex Libris 和英国版权许可代理机构(Copyright Licensing Agency, CLA)于近日宣布了他们在数字内容商店(Digital Content Store, DCS)URL API 上的合作。此次合作将有利于数字内容商店、Ex Libris Alma 资源管理解决方案和 Leganto 阅读清单解决方案之间的整合。Ex Libris 是致力于帮助 DCS URL API 发展的众多协作伙伴中的一员。

DCS 是一个为高等教育机构(Higher Education Institutions, HEIs)提供创新权利、技术和内容的平台。这一平台目前正由 CLA 和 HEIs、出版商, 以及技术伙伴协作开发。DCS 平台的核心包括一个可查询的 PDF 文档资源库(从 HEIs 数字化内容中复印或抽取), 以及一个工作流程管理工具。当一个机构在 CLA 的许可下将其内容数字化, PDF 文档将会被直接上传到 DCS 上, 之后即可用于分配给一个特定的课程。学生可以通过 DCS 上的一个安全链接访问文档内容。最为重要的是, 授课教师能够从中发现已经被外部机构上传的数字化内容, 并将这些内容分配到相应的课程。

有了新的 URL API, Alma 和 Leganto 将能够在 DCS 平台上检索 URL 指向的数字化内容, 使得研究人员和学生可以通过 Leganto 阅读列表访问相应的链接。通过充分利用其他 HEIs 已经数字化的内容, 图书馆可以大大节省为自己的用户进行内容数字化的时间和精力。

CLA 战略和数字部主任 Eela Devani 评论, “简化用户的版权结算流程是 CLA 现行使命的一部分。我们正与 Ex Libris 合作, 将 Leganto 和 Alma 解决方案与 DCS 整合在一起。通过整合 Leganto、Alma 和 CLA 新的权利和内容管理平台(DCS), 提供一个简化的工作流管理工具, 这一合作将为研究机构节省大量的时间和精力。”

Ex Libris Alma 产品管理部门主任 Asaf Kline 解释, “Alma 和 Leganto 与 DCS 之间的合作是我们之前做过的 CLA 的权限检查工具 API 工作的继续。此次整合将提供一个终端到终端的工作流程, 从 Leganto 的一个内容请求开始, 到 Alma 和 DCS API 的许可并提供材料, 直到将其传递给机构内学生为止。整个过程的执行将比我们想象的更加简单、便捷。”

(编译自: http://librarytechnology.org/news/pr.pl?id=21696)

(本刊讯)